

# Fallstudie: Einsatz von RealObjects PDFreactor® zur Erfüllung gesetzlicher Archivierungspflichten (PDF/A) in der Internet-Filiale der Sparkassen

## Über die Finanz Informatik

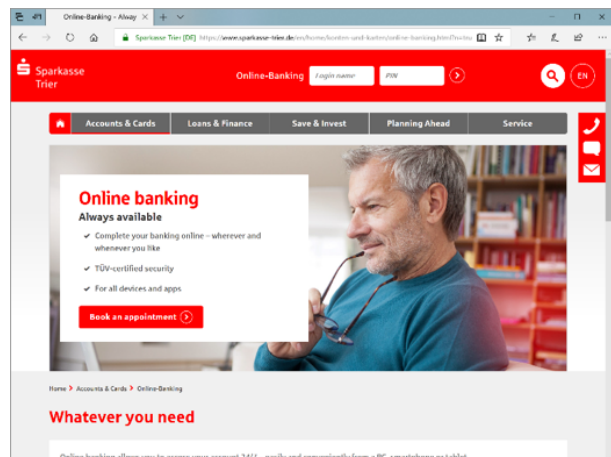
Die Finanz Informatik mit Sitz in Frankfurt am Main ist der zentrale IT-Dienstleister der Sparkassen-Finanzgruppe. Zu ihren Kunden gehören 384 Sparkassen, sechs Landesbanken, die DekaBank, acht Landesbausparkassen, öffentliche Versicherer sowie weitere Unternehmen der Sparkassen-Finanzgruppe und der Finanzbranche. Die Finanz Informatik bietet kompletten IT-Service – von Anwendungsentwicklung über Infrastruktur- und Rechenzentrumsbetrieb bis hin zu Beratung, Schulung und Support. Mit OSPlus stellt das Unternehmen heute eines der führenden Gesamtbankensysteme für den deutschen Markt.

Die Finanz Informatik übernimmt den Service für 117 Millionen Bankkonten; auf den Rechnern und Systemen werden jährlich mehr als 128 Milliarden technischer Transaktionen durchgeführt. Das Unternehmen beschäftigte Ende 2018 3.560 Mitarbeiterinnen und Mitarbeiter, die Umsatzerlöse betragen zusammen mit den Tochtergesellschaften rund 1,9 Milliarden Euro im Geschäftsjahr 2018.

Tochterunternehmen wie die Finanz Informatik Technologie Service, die Finanz Informatik Solutions Plus, die inasys und nicht zuletzt die Star Finanz ergänzen mit ihren Leistungsangeboten das IT-Portfolio der Finanz Informatik.

## Problemstellung

Laut Richtlinie 2004/39/EG des Europäischen Parlaments und des Rates vom 21.04.2004 über Märkte für Finanzinstrumente (MiFID, Abk. für Markets in Financial Instruments Directive) sind Banken dazu verpflichtet, ihren Kunden präsentierte Werbeinhalte für bestimmte Produktarten über einen Zeitraum von fünf Jahren zu archivieren und im Falle von Streitigkeiten vorzulegen. Dies betrifft auch online über die Internet-Filiale der Sparkassen angezeigte Werbeinformationen. Die Umsetzung einer richtlinienkonformen Archivierung (PDF/A konform) stellte die Finanz Informatik als Betreiber der Internet-Filiale vor erhebliche Herausforderungen.



Die Internet-Filiale der Sparkassen ist als Online-Banking-Portal eine komplexe und hochdynamische Web-Anwendung und keine Ansammlung statischer Seiten. Sie ist für die Sparkassen zentrale Drehscheibe für Service und Dialog mit ihren Kunden im digitalen Multikanal. Daher sind alle für den Endkunden relevanten Produkt-, Beratungs- und Serviceangebote in die Internet-Filiale integriert und werden permanent überprüft und bei Bedarf weiterentwickelt. Monatlich rund eine Milliarde Mal werden die Inhalts- und Banking-Seiten der Sparkassen aufgerufen – so oft, wie keine andere deutsche Webpräsenz. Die Sparkassen rangieren damit noch vor Seiten wie Bild.de, T-Online.de oder eBay Deutschland.

Laut MiFID können archivierungspflichtige Inhalte potenziell auf verschiedensten Seiten und Seitenbereichen sowie in verschiedensten Kontexten angezeigt werden. Die knapp 390 Sparkassen verfügen außerdem im Rahmen der Internet-Filiale über redaktionelle Freiheiten, die für weitreichende institutsspezifische Anpassungen und Erweiterungen zentral bereitgestellter Inhalte genutzt werden. Sowohl zentraler als auch institutsspezifischer Content wird häufig, teilweise täglich, aktualisiert. Darüber hinaus unterliegt der Programmcode der Anwendung einer ständigen Weiterentwicklung mit einem in der Regel halbjährlichen Release-Zyklus.

Die Finanz Informatik hat verschiedene Ansätze untersucht, wie sich unter diesen Voraussetzungen die MiFID-Anforderungen erfüllen lassen. Im Fokus stand auch die gesetzeskonforme Archivierung durch Unterstützung des Langzeitarchivierungsformats PDF/A, dem Format zur Langzeitarchivierung digitaler Dokumente, welches von der International Organization for Standardization (ISO) als Subset des Portable Document Format (PDF) genormt wurde.

### **Untersuchte Lösungsansätze ohne PDFreactor**

Ein naheliegender Ansatz wäre es, im Falle einer MiFID-Anfrage, d.h. auf Verlangen eine archivierte Werbeinformation vorzulegen, um die fragliche dynamische Seite inhaltlich exakt so zu reproduzieren, wie sie zum angefragten Zeitpunkt gültig war. Die Seite müsste dazu auf Basis des archivierten Contents und mittels des zum gegebenen Zeitpunkt gültigen Code-Standes der Anwendung neu generiert werden. Dabei würde es die ständige Weiterentwicklung des Anwendungscodes erforderlich machen, jeden jemals produktiv eingesetzten Code-Stand lauffähig auf einem System vorzuhalten oder zum Zeitpunkt der MiFID-Anfrage erneut zu installieren. Ein solcher Ansatz erwies sich bei der Komplexität der Anwendung und unter Berücksichtigung der eingesetzten Systeme und Frameworks als sehr aufwendig in der Realisierung.

Als weiterer Ansatz wurde geprüft, ob die Anforderungen mit dem im eingesetzten Redaktionssystem vorhandenen Mechanismus zur Archivierung der Redaktionsansicht von Seiten erfüllt werden können. Auch dies erwies sich als ungeeignet, da die Performance der Redaktionsanwendung bei einer ständig vorgehaltenen Datenmenge für rückwirkend fünf Jahre zu stark sinkt. Darüber hinaus wird bereits für die Erzeugung der Redaktionsansicht einer der Weiterentwicklung unterliegende Anwendungscode benötigt und es ergäbe sich das gleiche Problem wie im ersten Ansatz.

### **Realisierter Ansatz mit RealObjects PDFReactor**

Als letzter Ansatz wurde ein Verfahren untersucht, bei dem nach jeder redaktionellen Änderung des Contents oder einer Weiterentwicklung des Anwendungscodes eine statische Repräsentation für jede dynamisch generierbare und für MiFID relevante Seite erzeugt und archiviert wird. Damit ist keine dynamische Neugenerierung der Seiten zum Zeitpunkt einer späteren MiFID-Anfrage erforderlich und die oben genannten Probleme bei der Bereitstellung lauffähiger historischer Code-Stände oder großen, im produktiven System zu verwaltenden, Content-Mengen entfallen. Dennoch stellten sich auch bei diesem Ansatz schwierige technische Herausforderungen. Diese ließen sich aber durch den Einsatz des Produkts PDFReactor der Firma RealObjects sehr gut bewältigen. Dadurch war dieser Ansatz gegenüber den untersuchten Alternativen deutlich einfacher und kostengünstiger realisierbar und wurde damit zur Umsetzung ausgewählt.

Da die Internet-Filiale über ein Gruppenkonzept (50+ Gruppen, 400+ physikalische und/oder virtuelle Server) betrieben wird, entschied sich die Finanz Informatik für eine PDFReactor Corporate Application Lizenz, die auch den zukünftigen Ausbau der Server-Infrastruktur abdeckt. Welche technischen Herausforderungen bei diesem Ansatz bestanden und wie diese mittels PDFReactor gelöst werden konnten, wird im Folgenden näher betrachtet.

## Technische Details

Eine grundsätzliche Herausforderung bei diesem Ansatz besteht darin, eine geeignete statische Repräsentation der dynamisch erzeugten Seiten zu finden. Eine reine Textrepräsentation ist nicht ausreichend, da MiFID-relevante Inhalte auch in Bildern (z. B. werbewirksam präsentierte Zinssatz-Angaben) enthalten sein können. Auch räumliche Anordnung und Zusammenspiel von Text und Bild auf der Seite können für den transportierten Inhalt und seine MiFID-Relevanz von Bedeutung sein. Damit muss die archivierte statische Repräsentation der Seite visuell möglichst genau der vom Browser gerenderten Originaldarstellung entsprechen. Als statische Repräsentation kommen daher PDF-Dokumente in Frage. Allerdings besteht die Herausforderung dann genau in der Erzeugung eines visuell der Browser-Darstellung entsprechenden PDF-Dokumentes.

Diese Herausforderung lässt sich durch die speziellen Features der Code-Bibliothek PDFreactor, die sie von vielen anderen Bibliotheken zur PDF-Erzeugung abhebt, besonders gut bewältigen. Basis bildet ein PDF-Service in jeder der Server-Gruppen. Als wesentlichstes Feature übernimmt PDFreactor das Rendern des PDF-Dokuments (PDF/A 1b) direkt aus der HTML-Repräsentation der zu archivierenden Seite selbst. Im Gegensatz zu vielen anderen PDF-Bibliotheken müssen hier also vom Programmierer nicht Texte und Bilder per Code-Aufruf selbst in das zu erzeugende PDF eingefügt und dabei die Positionen, Farben und Größen explizit angegeben werden. Es reicht vielmehr aus, die URL der zu archivierenden Seite zu übergeben. PDFreactor arbeitet dann wie ein Browser, d.h. es führt die erforderlichen Requests zum Server, inklusive der Requests für eingebettete Ressourcen wie Bilder und CSS, aus, rendert die visuelle Darstellung aus der HTML-Repräsentation und gibt sie als PDF aus. Das heißt jede dynamisch generierte Webseite, die über eine URL abrufbar ist, kann mit PDFreactor sofort und ohne weitere selbst zu programmierende Zwischenschritte durch einen einzelnen Bibliotheks-Aufruf in ein archivierbares PDF überführt werden. Nur durch dieses Key-Feature von PDFreactor, wird der gewählte Ansatz zur Erfüllung der MiFID-Anforderungen in der Internet-Filiale so kostengünstig realisierbar.

Eine weitere Herausforderung besteht in der korrekten Umsetzung komplexer und auf Responsivität ausgelegter CSS-Layouts, wie sie in der Internet-Filiale zum Einsatz kommen, in der erzeugten PDF-Repräsentation. Auch diese Herausforderung ließ sich durch die hohe Qualität und Standardkonformität, mit der PDFreactor CSS-Formatierungen berücksichtigt, ohne Probleme meistern.

In der Internet-Filiale kommen darüber hinaus Javascript-Frameworks, wie z.B. JQuery und Plugins, zum Einsatz, die nach dem Laden einer HTML-Seite im Browser noch Formatierungen per Javascript vornehmen. Um diese Layouts korrekt im erzeugten PDF zu berücksichtigen, muss Javascript ausgeführt werden, wie das auch in einem Browser der Fall wäre. Auch dieses Feature bietet PDFreactor. Tatsächlich dynamische Elemente, wie z.B. ein rollierender Opener, können natürlich nicht in einem statischen PDF dargestellt werden. Dies ist aber Prinzip bedingt und hängt nicht mit einer Einschränkung von PDFreactor zusammen. Im Projekt wurde diese Schwäche des Ansatzes auf einfache Art überwunden, indem dynamische Elemente bei der MiFID-Archivierung in einer alternativen statischen Darstellung dargestellt werden. So werden beispielsweise alle Ausprägungen eines rollierenden Openers untereinander dargestellt.

Die Integration der Archivierungsfunktionalität in die Gesamtanwendung ist eine weitere Herausforderung. Alternative PDF-Bibliotheken, die eine Browser-Umgebung voraussetzen oder auf PHP oder anderen Scriptsprachen basieren, hätten hier unter Umständen Probleme bereiten können, da die Zielumgebung auf Java und OSGI basiert. Da PDFreactor aber als OSGI-Bundle bereitsteht, war eine einfache und stabile Integration gewährleistet.

## Fazit

Durch den Einsatz von RealObjects PDFreactor konnte für die Internet-Filiale ein Ansatz zur Gewährleistung gesetzlicher Archivierungsanforderungen (PDF/A konform) umgesetzt werden, der viel Aufwand gegenüber untersuchten Alternativansätzen gespart hat und der mit anderen PDF-Bibliotheken nicht ohne weiteres realisierbar gewesen wäre. Ein Finanz Informatik Entwickler merkte weiterhin an: *„Aus Entwicklersicht eine Top-Bibliothek mit enormen Leistungsumfang und sehr guter Integrierbarkeit. Auch die hohe Qualität des Supports ist hervorzuheben.“*

Finanz Informatik  
Theodor-Heuss-Allee 90  
60486 Frankfurt a. M.  
E-Mail: [info@f-i.de](mailto:info@f-i.de)  
[www.f-i.de](http://www.f-i.de)